

Ambika Timilsena

Data Immersion Portfolio
CareerFoundry
Airbnb Berlin Analysis



Introduction



Overview:

Airbnb is an American company that operates an online marketplace for lodging, primarily homestays for vacation rentals, and tourism activities. The platform is accessible via website and mobile app. Airbnb does not own any of the listed properties; instead, it profits by receiving commission from each booking.

Objective:

To analyze what variables may impact the price and demand of Airbnb Listings, which districts in Berlin are most popular among tourists and search for any noticeable patterns and trends in the data.

This project was created as a part of CareerFoundry Data Analytics Program

Role:

Data Analyst

Primary Stakeholder:

CareerFoundry Data Analytics Course

Tools Used:

- ★ Python
- ★ Tableau

Data Source

This public dataset is part of Airbnb, and the original source can be found on [Insideairbnb](https://insideairbnb.com)[1].

Python Skills

- ★ Linear Regression
- ★ K-Means Clustering
- ★ Time-series analysis
- ★ Seasonal Decomposition
- ★ Stationary Test
 - ❖ Dickey-Fuller Test
 - ❖ Autocorrelation Test
- ★ Geo- Visualization with Choropleth

Sources:

[1] <http://insideairbnb.com/get-the-data/>



Data Preparation

16,416 Rows
18 Columns

13,533 Rows
15 Columns

13,533 Rows
18 Columns

13,533 Rows
20 Columns

12,600 Rows
20 Columns

12,600 Rows
20 Columns

Original Data
Set

Consistency
Check

Combined
with Other
Required
Features

Derived New
Features

Removal of
Unwanted
Outliers

Final data
Set

1

2

3

4

5

6

Questions to Explore

01

Types of Room and Most Popular Location

Which types of room are available in Airbnb Berlin?

Which location is most popular for booking Airbnb in Berlin?

02

Variables Affecting Price

What factors affect the price of listings on Airbnb in Berlin?

Variables Affecting Demand

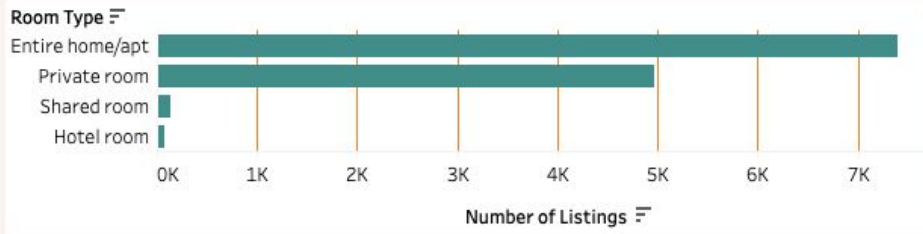
03

What factors affect the demand of listings on Airbnb in Berlin?



01 Types of Room in Airbnb

Types of Room Available in Airbnb Berlin



58.78%
Entire home/apartment

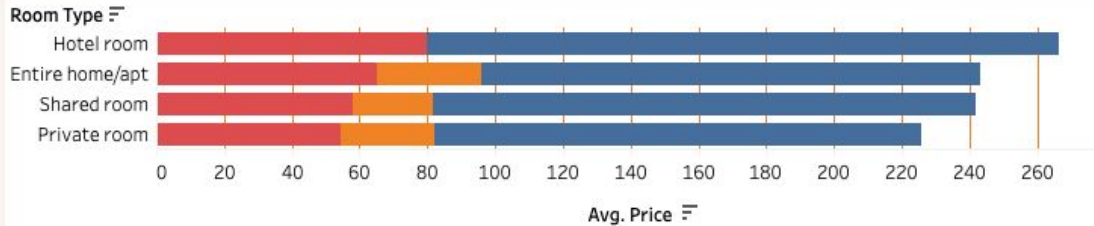
39.47%
Private room

1.09%
Shared Room

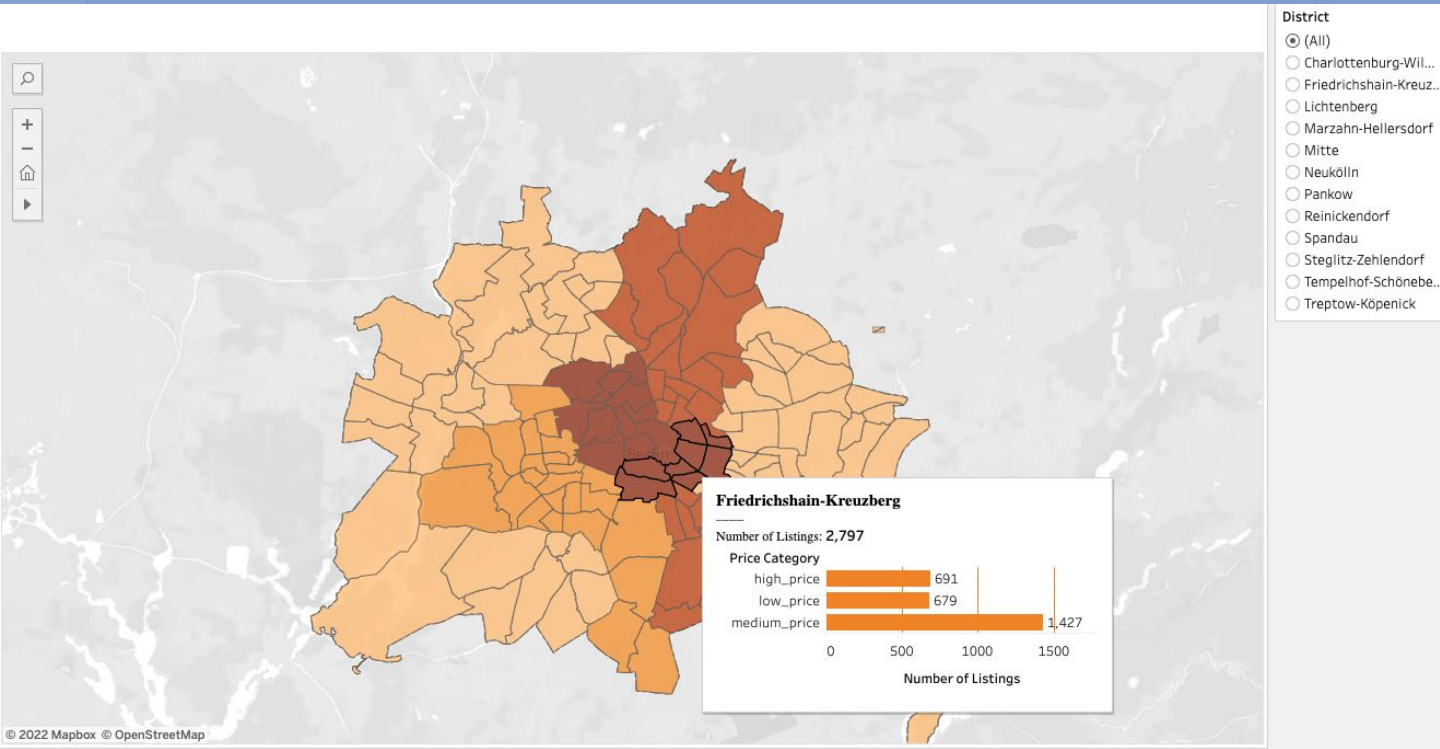
0.69%
Hotel Room

- Hotel rooms are the expensive room types in Airbnb with a very low share of rooms.

- Shared room has the lowest average price for booking.



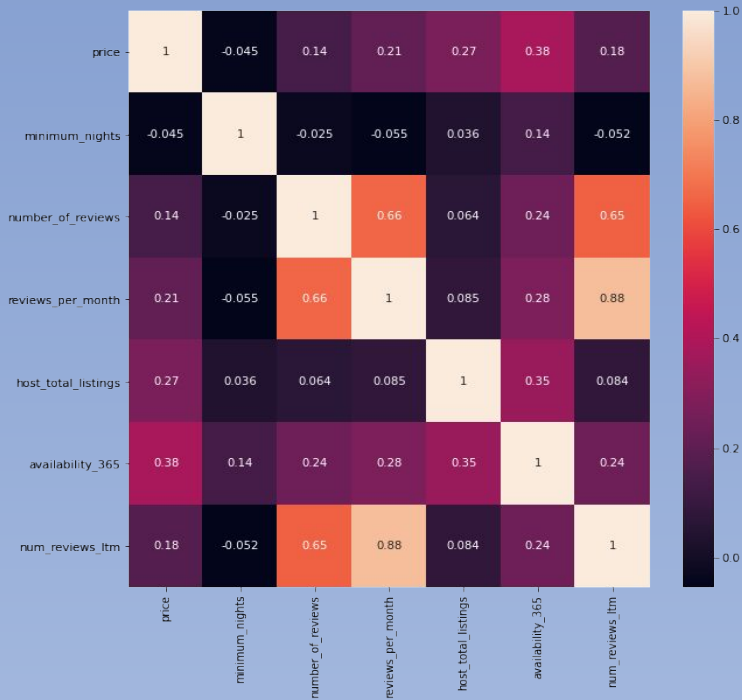
Most Popular Location



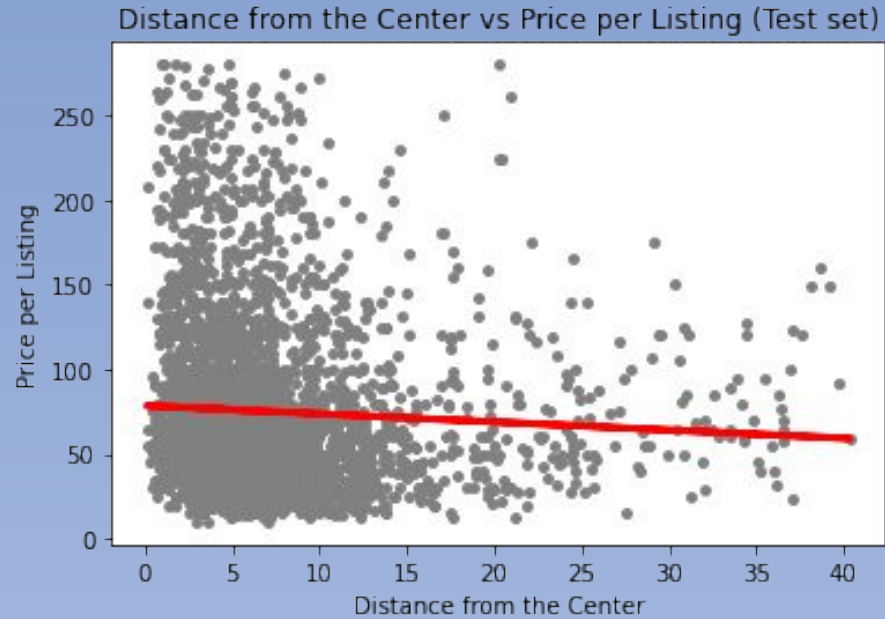
I considered the district with highest number of listings as most popular location.

Districts like **Mitte** and **Friedrichshain-Kreuzberg** have the highest number of listings. These districts are located at the **center of Berlin**.

02 variables Affecting Price



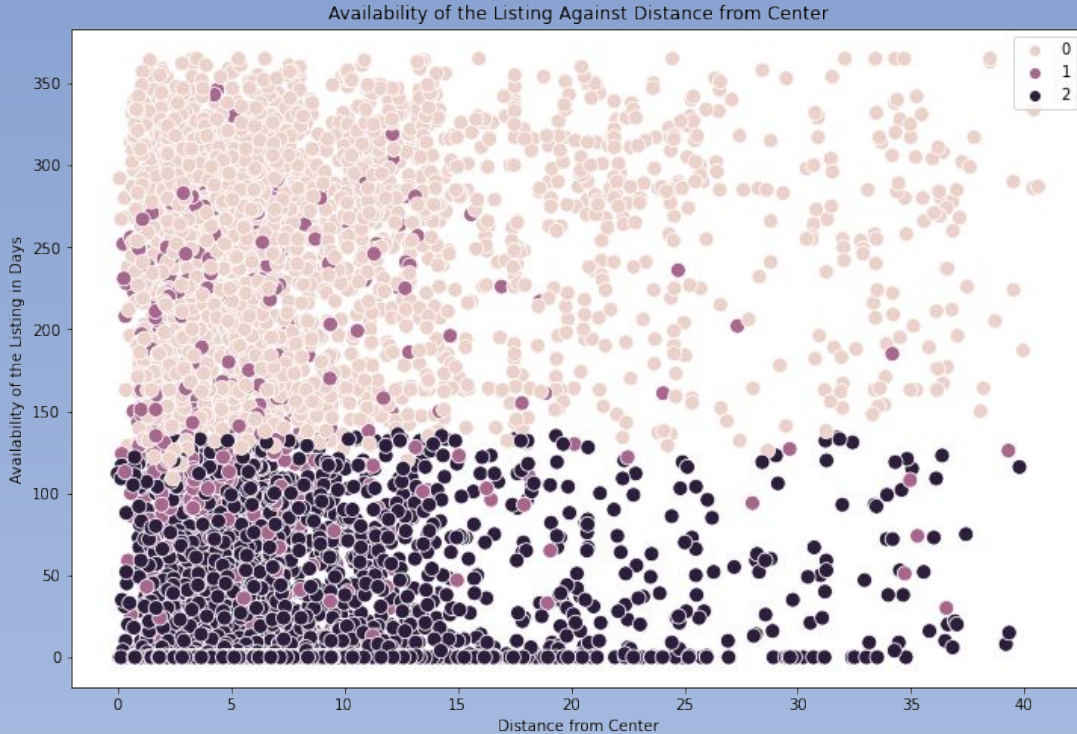
I used **correlation heatmap** to find correlation between variables affecting price of Airbnb Listings.



I trained uni-variate regression model using Distance from the center of Berlin to predict price of the Listing.

Result: Prediction using single variable not sufficient to predict price. Very Low R-Squared : 0.003

03 Variables Affecting Demand



I considered a listing as highly in demand if the **availability of listing** is closer to 0 days for next 365 days.

I used **K-Means Clustering using Elbow Curve** to find if demand is affected by distance from the center.

Result: Clustering based on single variable not sufficient.

Challenges

Looking back at the analysis stage, I realized that the biggest challenge was to find the variables which were directly affecting the price and demand of the Listings. My initial assumption was if the distance from the center of Berlin increases, price/demand decreases. However, after the exploration of data, I came to realize that price of the Airbnb Listings is not directly affected by distance, reviews or availability of the listings.

The data was limiting in size, only available for 2022. Further, there was no time series data of daily bookings for further analysis.



Recommendation

- ✔ In the end, the price and demand of the listings are dependent on several variables and it cannot be assessed or determined by only examining one of them. There are variables that contribute more towards the price of the listing; such as room type and the number of rooms.
- ✔ Friedrichshain-Kreuzberg and Mitte have the highest number of listings and are most popular district for Airbnb Booking.
- ✔ There are certain areas/districts in Berlin which have high price-range listings like Mitte, Pankow, and Spandau.
- ✔ We can find listings with the lowest prices in Reinikendorf and Neukölln District.

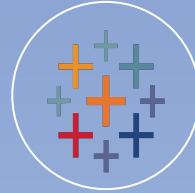


Do You Have Any Questions?

Here are additional links with more recommendations and python code:



[Github Link](#)



[Data Visualization in Tableau](#)